University of Southampton

An Empirical Review of Uncertainty Estimation for Quality Control in CAD Model Segmentation

Gerico Vidanes, David Toal, Andy Keane, Daniel Xu Zhang, Marco Nunez, Jon Gregory

26 June 2025

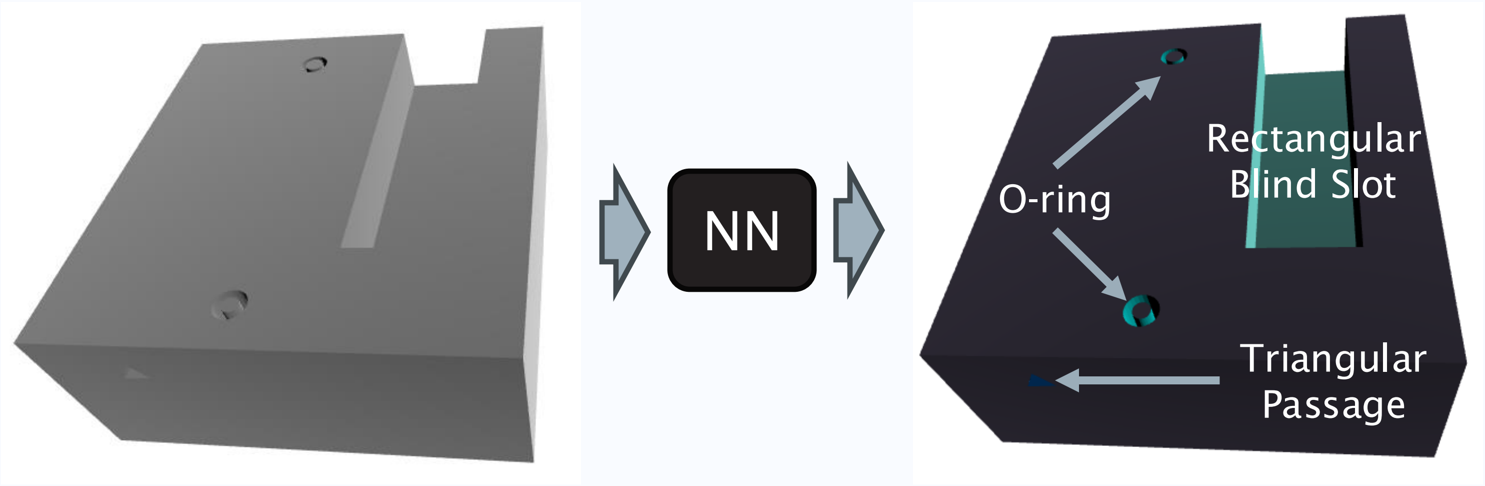
Overview

- Context
- Background
- Practical Uncertainty Estimation Methods
- Evaluation Method
- Results
- Conclusions

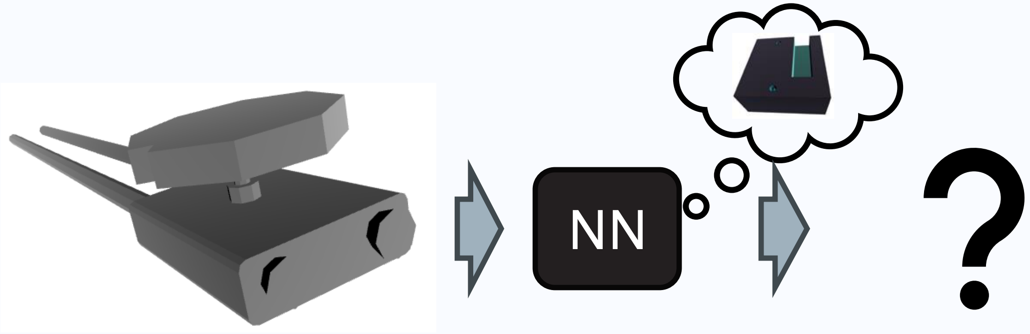
Context

Quality control and confidence of CAD model segmentations

- Semantic segmentation of CAD models enables wide range of automations.
  - CAD to CAM
  - CAD to Simulation



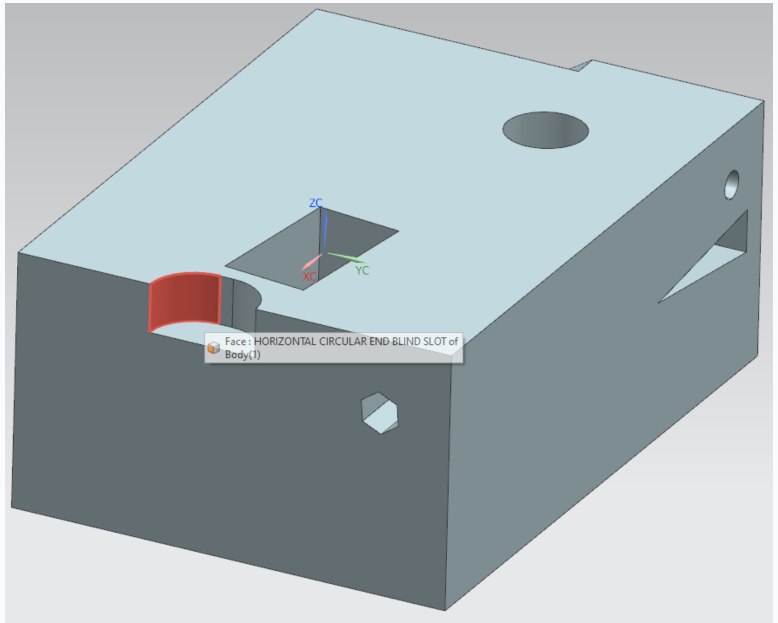
- General feature recognition on geometries is widely useful in engineering for automated decision making and processing.
- Modern NN approaches can achieve high accuracy **but not perfect.** A signal for uncertainty useful for:
  - Flagging potentially incorrect predictions.
  - Unfamiliar inputs



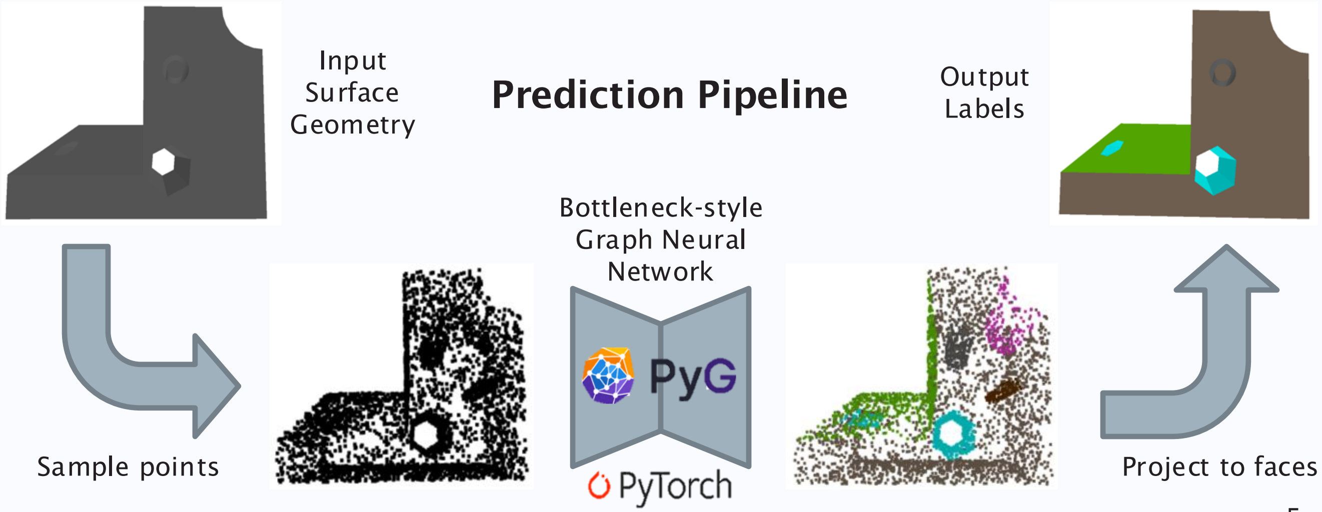
Background

B-Rep CAD model segmentation with neural networks

- B-rep CAD models ~= collection of surface patches.
- Using point-based NN [1].
  - Outputs a vector for each b-rep face.
  - Softmax as probability? **This is the baseline:** $\max_j\left(\frac{e^{z}}{\sum_j e^{z_j}}\right)$



- Standard training
  - Adam optimizer
  - Cross entropy loss
  - 'Best' model from lowest loss on validation set.



[1] Vidanes, G., et al. CAD (2024) - *Extending Point-Based Deep Learning Approaches for Better Semantic Segmentation in CAD*

Practical Uncertainty Estimation Methods

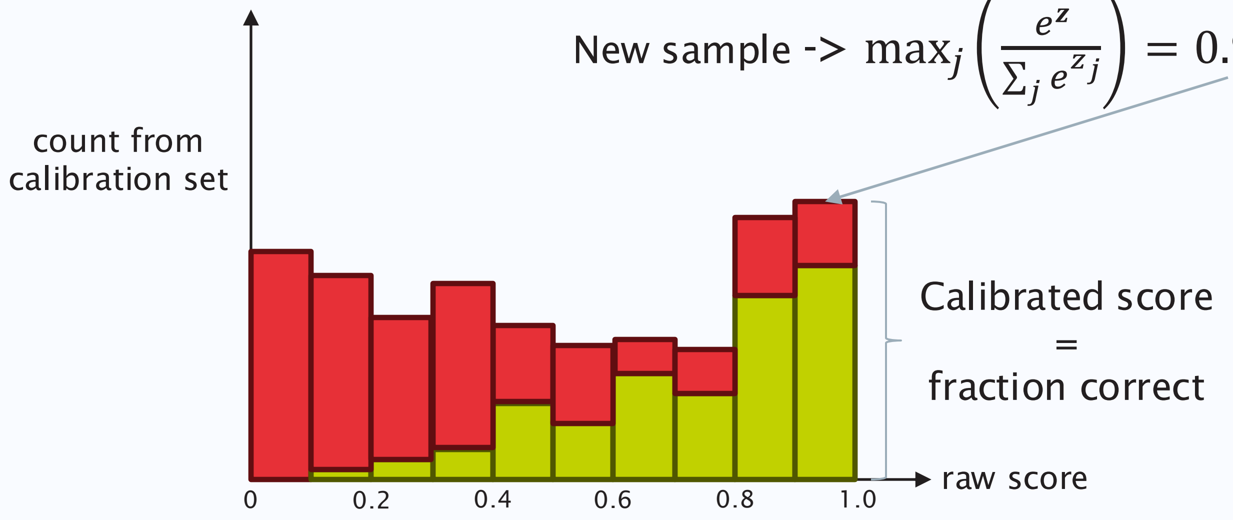
- Post-processing calibration
  - Temperature scaling [1]
  - Histogram binning [2]

$$P(\hat{\boldsymbol{y}}) = \frac{e^{\boldsymbol{z}/T}}{\sum_j e^{z_j/T}}$$

Labelled Calibration Dataset

Find T which minimises cross entropy loss given a <u>trained</u> NN

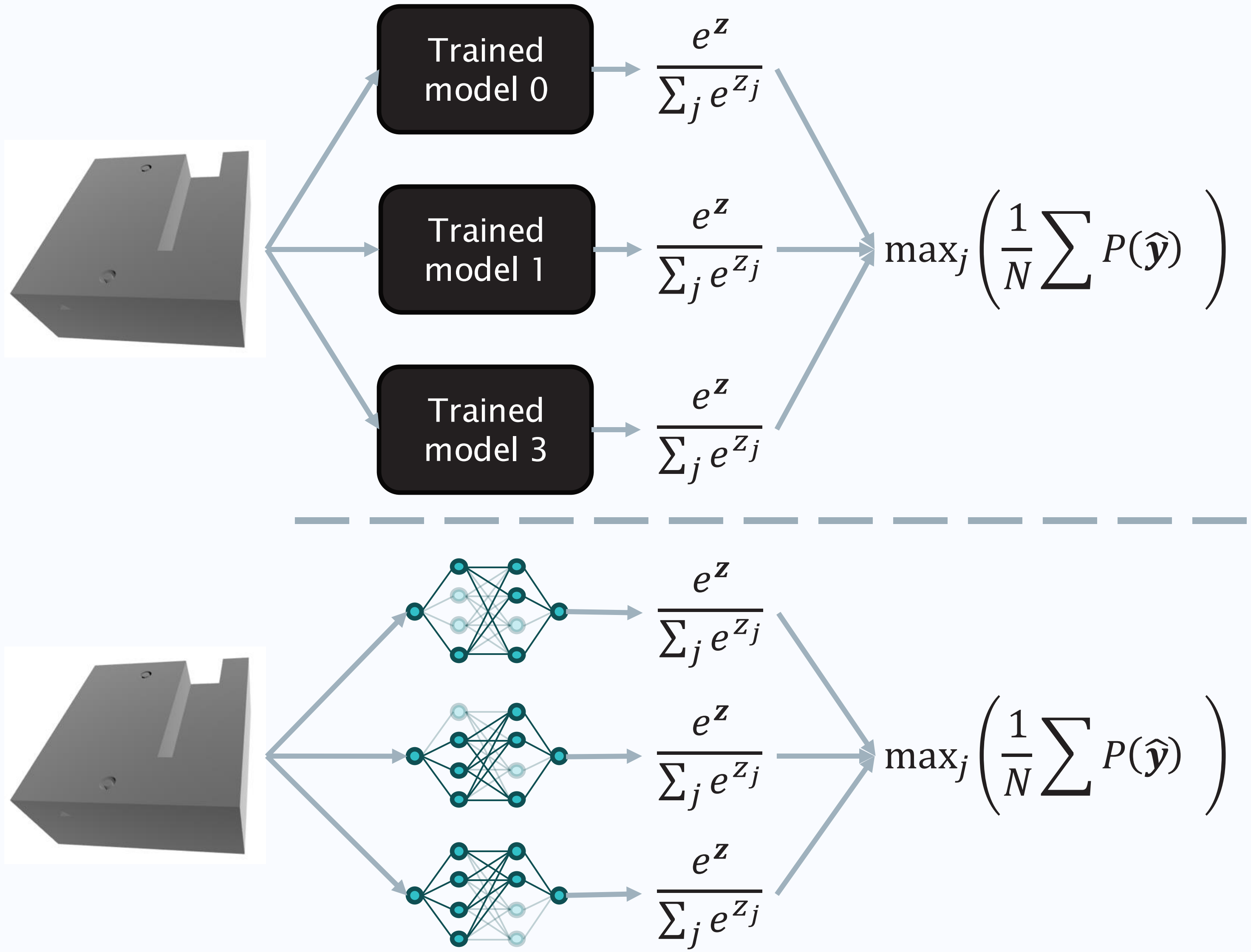
New sample -> $\max_j \left(\frac{e^{\boldsymbol{z}}}{\sum_j e^{z_j}} \right) = 0.92$



[1] Guo, C., et al. ICML (2017) - *On Calibration of Modern Neural Networks*
[2] Zadrozny, B. and Elkan, C. ICML (2001) - *Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers*

Practical Uncertainty Estimation Methods

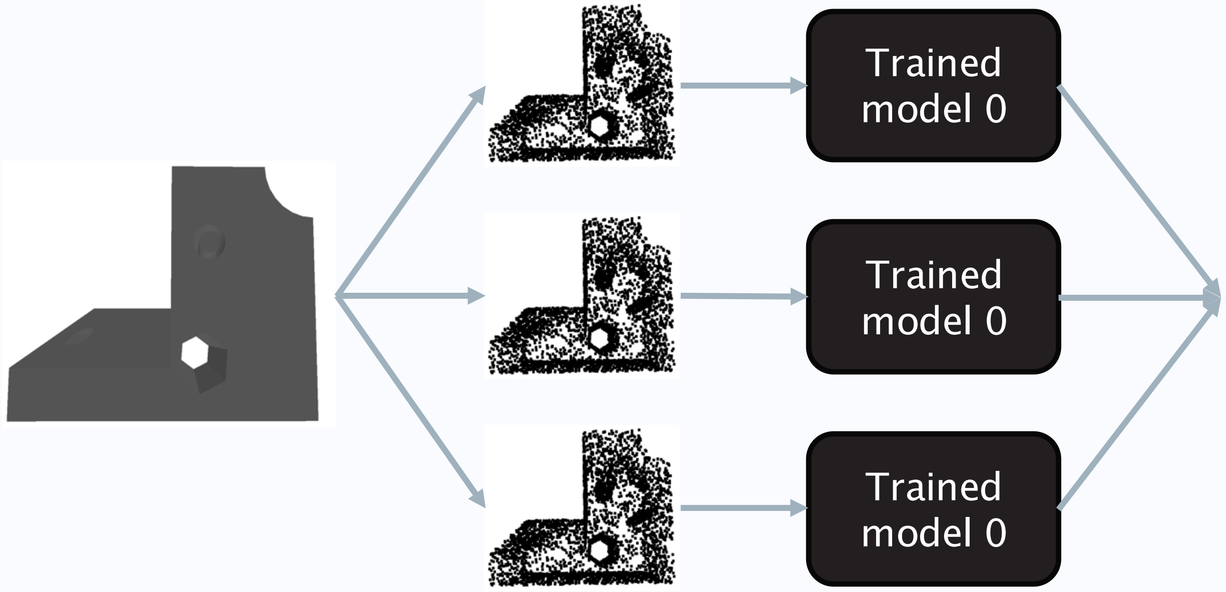
- Post-processing calibration
  - Temperature scaling
  - Histogram binning
- Model stochasticity
  - Deep Ensemble [1]
  - MC Dropout [2]



[1] Lakshminarayanan, B., et al. NeurIPS (2017) - *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*
[2] Gal, Y. and Ghahramani, Z. PMLR (2016) - *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*

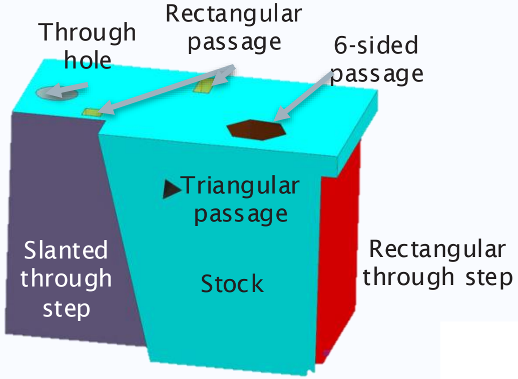
Practical Uncertainty Estimation Methods

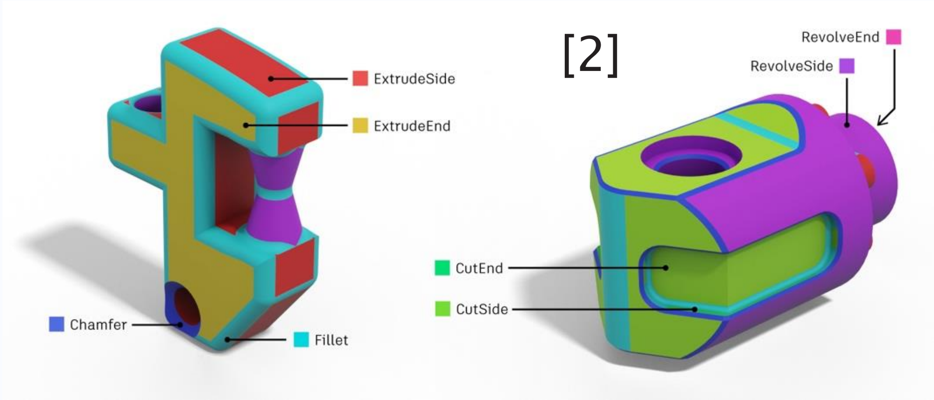
- Post-processing calibration
  - Temperature scaling
  - Histogram binning
- Model stochasticity
  - Deep Ensemble
  - MC Dropout
- Input data augmentation
  - Point resampling



$$\max_j \left(\frac{1}{N} \sum P(\hat{\boldsymbol{y}}) \right)$$

Evaluation Method





- Benchmark datasets/cases:
  1. MFCAD++ [1] – Trained on ~43k geometries
  2. MFCAD++ (5%) – Trained on ~2k geometries
  3. Fusion360 Gallery [2] – Trained on ~26k geometries

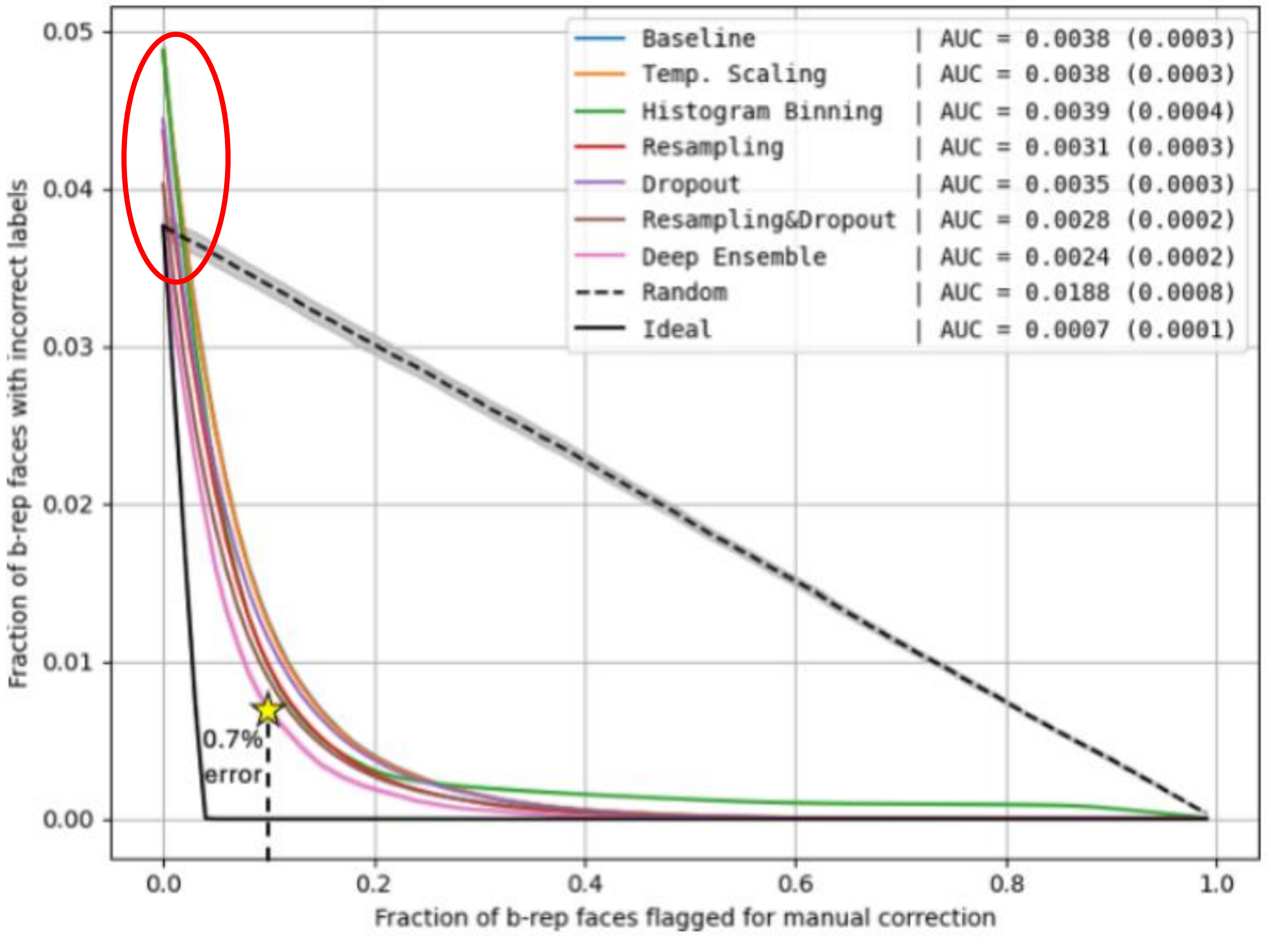
1000 unseen geometries from relevant dataset → Trained model(s) → B-rep face predictions → Uncertainty approach → Uncertainty/Confidence scalar for each prediction

- Quality control scenario adapted from [3]
  - Predictions ranked from most uncertain to least uncertain.
  - N predictions flagged for correction. Remaining error is measured.

[1] Colligan, A., et al. CAD (2022) - *Hierarchical cadnet: Learning from b-reps for machining feature recognition*
[2] Lambourne, J.G., et al. CVPR (2021) - *Brepnet: A topological message passing system for solid models*
[3] Ng, M., et al. IEEE Trans. Biomed Eng. (2023) - *Estimating Uncertainty in Neural Networks for Cardiac MRI Segmentation: A Benchmark Study*

Results

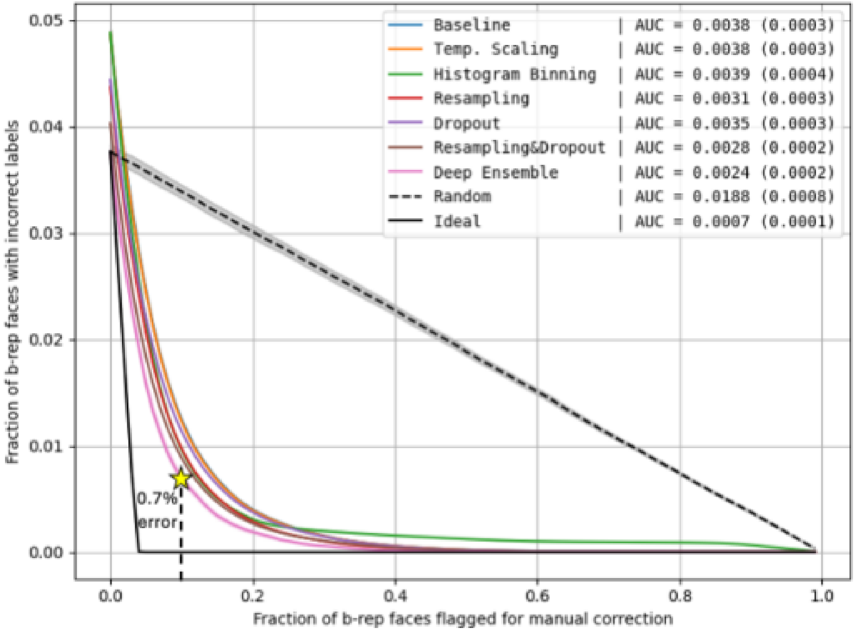
- Human-in-the-loop error, given manual correction budget.
  - Summarized by area-under-the-curve.
  - Lower is better
- Different error rates at 0% manually corrected.
  - Aggregation approaches can improve accuracy!



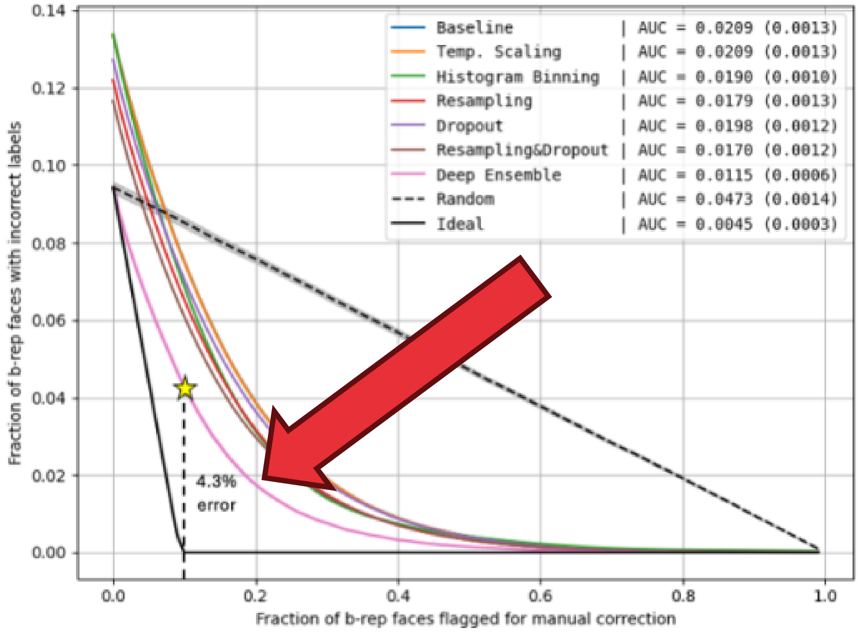
(a) MFCAD++

Results

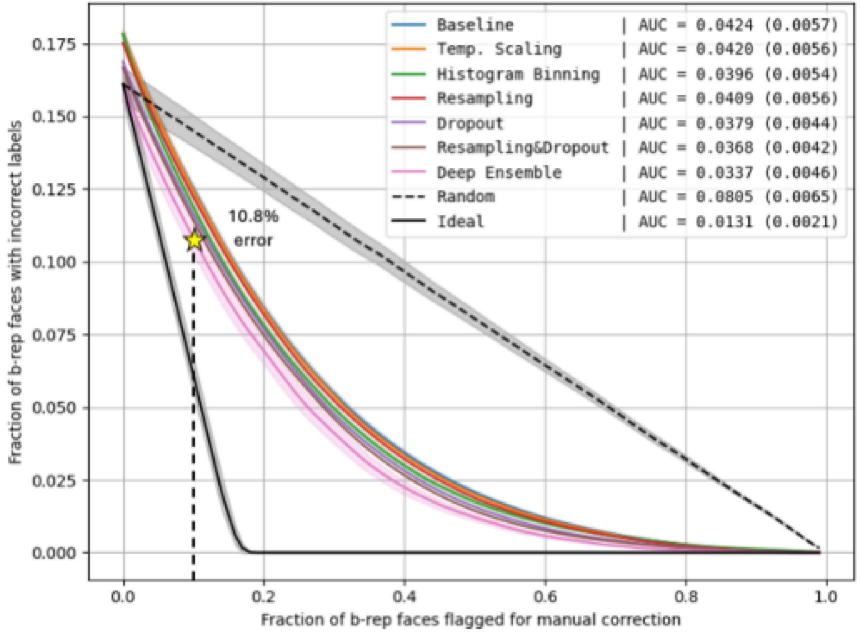
- Human-in-the-loop error, given manual correction budget.
  - Summarized by area-under-the-curve.
  - Lower is better
- Deep Ensemble particularly good at limited training data case!



(a) MFCAD++



(b) MFCAD++ (5%)



(c) Fusion360 Gallery

Trade-offs

More calibrated uncertainty →

Baseline, Temp. Scaling, Histogram Binning, MC Dropout, Resampling, Resampling & Dropout, Deep Ensemble

Less computational cost (inference only) →

Deep Ensemble, Resampling & Dropout, Resampling, MC Dropout, Histogram Binning, Temp. Scaling, Baseline

Less computational cost (total) →

Deep Ensemble, Temp. Scaling, Resampling & Dropout, Resampling, MC Dropout, Histogram Binning, Baseline

g.vidanes@soton.ac.uk

Conclusions

- For this setting, uncertainty estimation techniques tested are similarly good. Baseline is not as bad as expected!
- Relatively simple approaches are ‘sufficient’ for this point-based GNN for CAD segmentation.
- Human-in-the-loop approach validated for CAD segmentation. A signal is obtained which can rank predictions from most likely to least likely to be incorrect.
  - Opens up other uses for uncertainty…

An Empirical Review of Uncertainty Estimation for Quality Control in CAD Model Segmentation ⋆

Gerico Vidanes[1], David Toal[1], Andy Keane[1], Daniel Xu Zhang[2], Marco Nunez[3], and Jon Gregory[3]

[1] University of Southampton, UK
[2] Falmouth University, UK
[3] Rolls Royce plc., UK

Thanks to Rolls-Royce plc. for funding this work

YOUR QUESTIONS

g.vidanes@soton.ac.uk

More Results

Correlation of uncertainty with predictive accuracy

- Measured using conditional probabilities from [1].

1. $p(\textbf{accurate}|\textbf{certain})$: The probability that the model is accurate on its output given that it is confident on the same.
2. $p(\textbf{uncertain}|\textbf{inaccurate})$: The probability that the model is uncertain about its output given that it has made a mistake in its prediction (i.e., is inaccurate).

$$\mathbb{P}(\text{accurate}|\text{certain}) = \frac{n_{ac}}{n_{ac} + n_{ic}}$$

$$\mathbb{P}(\text{uncertain}|\text{inaccurate}) = \frac{n_{iu}}{n_{ic} + n_{iu}}$$

$$\text{PAvPU} = \frac{n_{ac} + n_{iu}}{n_{ac} + n + au + n_{ic} + n_{iu}}.$$

- Relevant frequencies/counts:
  - n_{ac} - accurate and certain, n_{au} - accurate and uncertain
  - n_{ic} - inaccurate and certain, n_{iu} - inaccurate and uncertain

[1] Mukhoti, J. and Gal, Y. arXiv (2018) - *Evaluating bayesian deep learning methods for semantic segmentation*

More Results

- Mean value of conditional probabilities using different uncertainty thresholds.
  - Higher is better.
- Similar trends as before.
  - But shows that using the uncertainty from Deep Ensemble gives lower recall (but more precise) filter.